



Hiroto Honda

Jul 25, 2018 · 5 min read · [Listen](#)

Introduction to deep super resolution

Hi, I'm Hiroto Honda, an R&D engineer at DeNA Co., Ltd, Japan. In this article I would like to introduce the recent progress on single-image super-resolution (SISR).

SISR aims at restoring a high-resolution image with rich details from a single low-resolution image. In recent years, SISR is getting more powerful and accurate thanks to the progress of convolutional neural networks (CNNs). I would like to show how you can start training SISR, progress of SISR network architectures from SRCNN to EDSR, and comparison of accuracy and runtime between the methods.

SISR is easy to try

To train an SISR network, firstly you have to prepare a dataset containing high-resolution (HR) and low-resolution (LR) image pairs. Costly manual annotations are not necessary — you just have to gather HR images and apply downsampling filter on them to get LR images. The CNN tries to learn the inverse function of the downsampling filter to restore the lost details of the HR images.

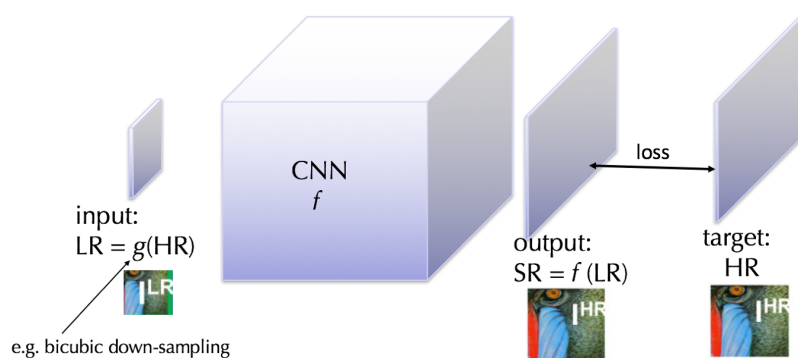


Fig. 1 How single-image super-resolution works (images from [4])

Here is how you can start training a deep SISR network (see Fig. 1).

1. Gather HR images.
2. Crop patches from the HR images. (e.g. 96×96)
3. Down-sample them to generate input images $LR = g(HR)$.
4. Put them into a batch $\{LR\}, \{HR\}$.
5. Train the network f with the pixel-wise loss function : e.g.

 Search


Hiroto Honda

260 Followers

$$\text{MSE}(\{\text{HR}\}, f(\{\text{LR}\})).$$

6. ...that's it!

For evaluation, Peak Signal-to-Noise Ratio (PSNR, in decibels) and Structural Similarity index (SSIM) are used. In this article we use PSNR for comparison between methods.

SRCNN, VDSR and ESPCN

The first CNN-based SISR method called SRCNN was introduced by Dong et al. in ECCV 2015 [1]. SRCNN consists of only three convolution layers, but outperforms the previous non-deep approaches. Very Deep Super Resolution (VDSR) [2] employs the similar structure as SRCNN, but goes deeper to achieve higher accuracy. Both SRCNN and VDSR apply bicubic upsampling at the input stage and deal with the feature maps at the same scale as output.

Shi et al. proposed Efficient Sub-Pixel Convolutional Neural Network (ESPCN) to make SRCNN more efficient [3]. ESPCN deals with the feature maps at LR resolution and upsampling is carried out afterwards, which makes the total amount of computation much smaller than SRCNN.

For upsampling sub-pixel convolution (combination of a convolution and a 'pixel shuffle' operation) is exploited. Pixel shuffle rearranges the elements of $H \times W \times C \cdot r^2$ tensor to form $rH \times rW \times C$ tensor (Fig. 3). The operation removes the handcrafted bicubic filter from the pipeline with little increase of computation.

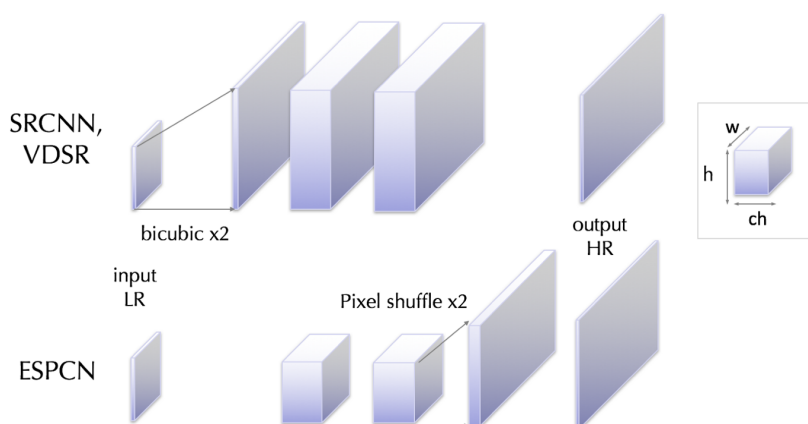
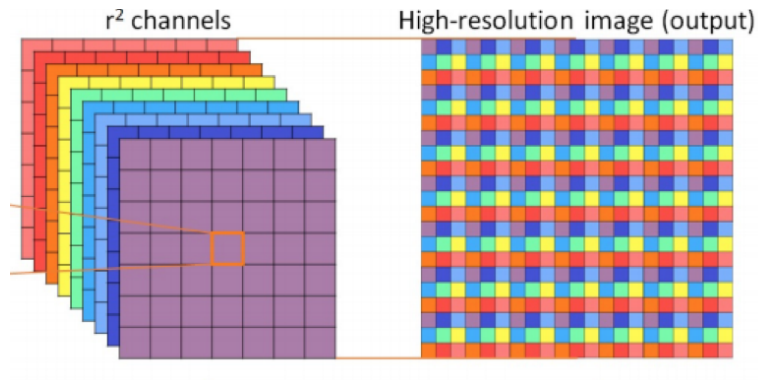


Fig.2 Difference between SRCNN, VDSR, and ESPCN.

[Help](#) [Status](#) [Writers](#) [Blog](#) [Careers](#) [Privacy](#) [Terms](#)
[About](#) [Knowable](#)

Fig. 3 Pixel Shuffle operation ($r = 3$). [3]

SRResNet

Ledig et al., introduced the stronger baseline called SRGAN at CVPR 2017 [4]. The network is largely based on ResNet architecture and introduces series of resblocks (Fig. 3). Unlike ResNet which down-samples the input up to $1/32$ scale, the resblocks of SRGAN operate on a single scale (input LR scale). The blocks 'modify' the input LR features gradually as they propagate deeper to get ready for the upsampling. For upsampling, pixel shuffle operator is used to avoid checkerboard artifacts [3].

In the paper three types of loss functions are proposed. 1) MSE loss 2) VGG (content) loss 3) adversarial loss. The network is called 'SRResNet' when only MSE loss is used.

SRResNet employs:

- 16 residual blocks with 64 channels
- Global skip connection
- pixel-wise L2 loss
- Pixel Shuffle upsampling

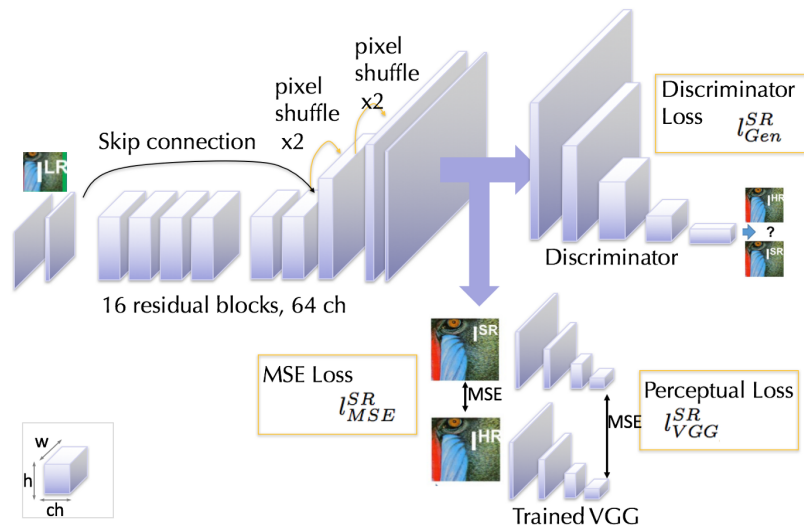


Fig. 3 SRGAN network structure [4]. It is referred to as 'SRResNet' when only MSE loss is used.

Enhanced Deep Residual Networks (EDSR)

In 2017 B. Lim et al. developed more advanced network called EDSR [5] and won the NTIRE 2017 Super-Resolution Challenge [6]. They started from SRResNet and optimized it for achieving further accuracy.

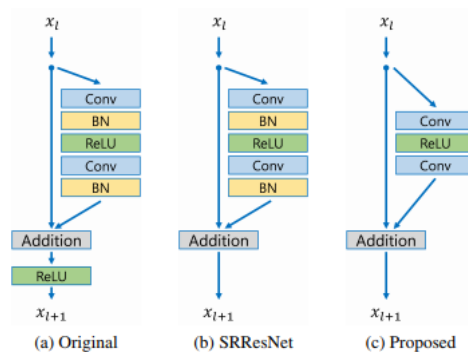


Fig. 4 Comparison of Resblock components between (a) original ResNet, (b) SRResNet, and (c) EDSR. [4]

EDSR network employs:

- 32 residual blocks with 256 channels
- pixel-wise L1 loss instead of L2
- no batch normalization layers to maintain *range flexibility* (Fig. 4)
- scaling factor of 0.1 for residual addition to stabilize training

Fig. 5 shows the comparison of PSNR and appearances between SISR methods. See how appearances change from 22.66 dB (Bicubic) to 23.89dB (EDSR). 1dB difference matters a lot!

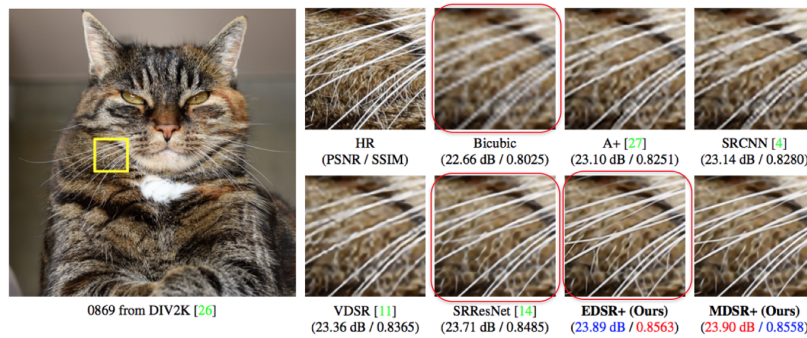


Fig. 5 PSNR and appearances [5]

Progress on SISR — accuracy and runtime

Let us summarize the progress of SISR methods. Fig. 6 shows the comparison of PSNR gain values (from [5]) over bicubic upsampling evaluated on Set5 dataset. Compared with SRCNN, SRResNet and EDSR achieve higher PSNR by 1.57dB and 2.14dB respectively. Meanwhile, as the network get deeper and wider (more channels), calculation gets more expensive. The bars shown in Fig. 5 represent number of mega-multiplication per one input pixel for x2 restoration. EDSR requires 30 times more multiplication than SRResNet, mainly due to larger number (4 times) of resblock channels.

We see a trade-off between performance and speed. ESPCN looks the most efficient while EDSR the most accurate but expensive. It depends on your application which method to choose.

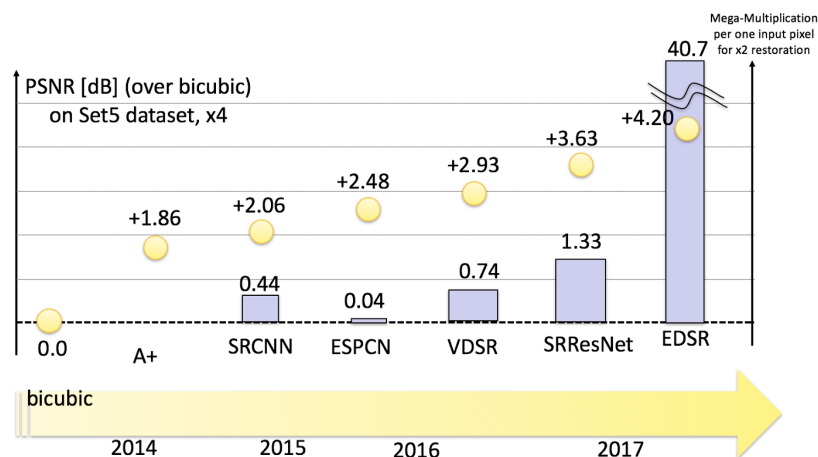


Fig. 6 Progress on SISR in aspect of accuracy (PSNR) [5] and runtime.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In ECCV, 2014.

- [2] J. Kim, J. K. Lee, and K. M. Lee. Accurate image superresolution using very deep convolutional networks. In CVPR, 2016.
- [3] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In CVPR, 2016.
- [4] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, 2017.
- [5] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In CVPRW, 2017. torch implementation: <https://github.com/LimBee/NTIRE2017> pytorch implementation: <https://github.com/thstkdgus35/EDSR-PyTorch>
- [6] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, et al. Ntire 2017 challenge on single image superresolution: Methods and results. In CVPR 2017 Workshops.